Dynamical Systems to Account for Turn-Taking in Spoken Interactions

Mathieu Jégou[1,2], Pierre Chevaillier[1], and Pierre De Loor[1]

1 ENIB–UEB; Lab-STICC; F29200 Brest, France
{pierre.chevaillier,pierre.deloor}@enib.fr
2 Technologic Research Institute b<>com ; F29200 Brest, France
mathieu.jegou@b-com.com

1 Introduction

Turn management is considered as essential for an Embodied Conversational Agent (ECA) to increase user's engagement with it [2]. This article presents a dynamical model for turn management in dyadic interactions. The model is a system of differential equations that mixes two models from the cognitive sciences, the Drift Diffusion Model, and the Behavioral Dynamics. Decision-making and the control of actions are two coupled processes that modulate continuously the behavior of the interacting agent. This conceptual model accounts for the emergence of smooth transitions without using neither prediction nor planning of the agent's behavior. The objective was not to obtain a fully realistic behavior, but to show how the model could account for the main qualitative properties of turn management, such as interrupting the current speaker, signaling its willingness to go on speaking, or yielding the turn to the next speaker.

2 Conceptual Model for Turn Management

2.1 Turn-Taking without Prediction

In their seminal work, Sacks et al. made a fundamental observation: participants exchange turns in a smooth way, most of the time without overlaps nor too long pauses [4]. To explain that, they proposed that listeners predict the end of a turn constructional unit to identify Transition Relevant Places (TRP). They do it by integrating a set of non verbal and verbal cues to identify when a TRP will occur [1]. Nevertheless the active role that listeners play in the emergence of turn transition [6] and the importance of signal variations for a transition to take place [1], make us claim that the occurrence of transitions is a self-organized, co-creative process, emerging from the interaction between participants. Based on some authors' works (see [1] for a review), we hypothesized that a conversational agent can rely mainly on non verbal signals to manage smooth turns.

2.2 Behavioral Architecture

Fig. 1 summarizes the principles of our behavioral architecture. First, the agent has an intrinsic motivation to be the speaker or the listener, depending on the conversational context. This communicative intention is under the control of the dialogue manager that generates some communicative intentions, captured here by the variable I (not controlled by our model). This intention depends on the conversational context (what the agent has to say), its personality or its mental state. Moreover, the agent will act to become the next speaker (or the next listener), or to keep its current role, depending on the non verbal cues it can get from the other participant's behavior. Acting means here producing verbal and non verbal signals. The loose coupling between the production of signals, the agent's own intention, and its perception of the other's behavior creates a complex relationship between the tendency to act on its own, and to be influenced by the other participant's behavior. As a result, turn management is emergent: no particular agent controls the occurrence of turn transitions, nor the duration of the transitions.

In our model, the agents continuously produce signals, following the principles of the behavioral dynamics elaborated by Warren [5]. In his view, behavior is self-organized, emerging from the interaction between the agent and its environment. The agent does not control entirely its behavior, but explores the global dynamics of the interaction, and adjusts its action to reach its goal. Besides, agents may have to make a decision (eg. to yield the turn or not) based on uncertain, if not contradicting, information about the intention of the opposite agent. The process of integrating evidence about the other agent's intention is controlled by the Drift Diffusion Model (DDM) [3]. The variable γ is the resulting confidence the agent holds about the intention of the other agent.

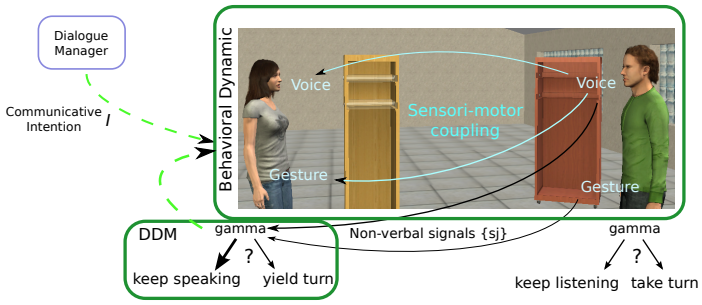


Fig. 1: Illustration of the role of the DDM and of the behavioral dynamic.

Decision-Making. The DDM accounts for human decision-making when an agent has to choose between two alternatives [3]. It assumes that agents continuously integrate over time the difference in the noisy information favoring each alternative they get and choose the most favorable alternative when this accumulated

value reaches a given threshold. We implemented the DDM as follows:

$$d\gamma = \alpha dt + \sigma d\epsilon \quad ; \quad \alpha(t) = \sum_{j=1}^{n_s} \alpha_j \Big(s_j(t), \dot{s_j}(t) \Big) \tag{1}$$

The model computes γ by integrating a set of signals $\{s_j\}$ produced by the other agent. It defines two thresholds $t_\gamma^+ = 1$, and $t_\gamma^- = -1$. When γ raises up t_γ^+, the agent considers that the other is willing to change its role, when γ falls down below t_γ^-, the agent considers that the other is not willing to change role. When γ is between the two thresholds, the agent is more or less confident about one or the other alternative. The drift coefficient α sums the accumulation of evidence corresponding to each signal j.

Sensory-motor coupling. For each non verbal signal s_j, the agent varies its production according to Warren's general equation:

$$\ddot{s_j} = -b\dot{s_j} - f_{\dot{s_j}}(s_j, \gamma, I) \tag{2}$$

The specific shapes of non verbal actions are defined by Eq. 3:

$$f_{\dot{s_j}}(s_j, \gamma, I) = \underbrace{t_\gamma^+ k_1 (s_j - c_1)}_{\text{a1}} + \underbrace{t_\gamma^- k_2 (s_j - c_2)}_{\text{a2}} + \underbrace{(1 - t_\gamma^+)(1 - t_\gamma^-) f'_{\dot{s_j}}(s_j, \gamma, I)}_{\text{a3}} \tag{3}$$

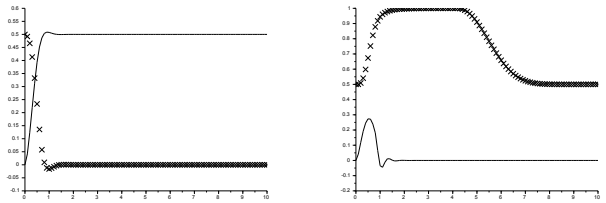
where: $t_\gamma^+ = 1$, if $\gamma \geq t_\gamma^+$ (0 otherwise) and $t_\gamma^- = 1$ if $\gamma \leq t_\gamma^-$ (0 otherwise).

Depending on the result of the accumulation process, one particular action, i.e. one of the terms `a1`, `a2` or `a3` is executed. `a1` is executed when the agent has accumulated enough evidence about option 1 ($\gamma = 1$), and `a2` is executed when evidence are against this option ($\gamma = -1$). `a3` is applied when $\gamma \in \left[t_\gamma^-, t_\gamma^+\right]$.

3 Results

The implementation of the model produced emerging turn transitions that satisfied the qualitative properties of human behavior.

$I \in [0, 1]$ denotes the agent's communicative intention: $I < .5$, the agent is not willing to change role, $I > .5$ the agent is willing to change role. Signals produced by the agent are: the intensity of the voice (V), the relative orientation of the agent to its interlocutor (B), and the arm gestures (G). Each signal reduces to one state variable, resp. s_v, s_b, s_g: the variation of each signal j along the time matches the one of a continuous variable $s_j \in [0, 1]$. The equations of the signals were devised to account for the following behaviors. V_S: the speaker lowers its voice to yield the turn, and speaks louder when it does not want to yield the floor whilst the other wants to take it. V_L: The agent starts speaking when it is confident about the willingness of the current speaker to yield the turn, G: the listener makes gestures to indicate it wants to take the turn, B: the higher the agent's intention to change role, the faster it faces its interlocutor.



(a) $I_S = 1.$; $I_L = 1.$ (b) $I_S = .0$; $I_L = 1.$

I_S: speaker's intention to give the turn; I_L: listener's intention to take it.

Fig. 2: Time series of the voice intensity of the Speaker V_S (crosses) and the Listener V_L (plain) in two scenarios.

Different scenarios, corresponding to different communicative intentions, have been simulated. Fig. 2 shows two examples of agent-agent interactions. It shows that the model reproduces different patterns depending on the scenario: Fig. 2a, shows two agents that are strongly willing to change turns and Fig 2b, a speaker that strongly wants to keep the floor and a listener that strongly wants to take the turn. In the first case, the turn actually occurs, not in the second one.

4 Conclusion

Simulations of agent-agent interactions show that our model reproduces the richness of turn management behavior. This is the first step towards a realistic agent. Our next goal is now to define the equations that could produce realistic behaviors, and to evaluate the realism of our agent by confronting it to users.

Acknowledgments. This work was supported in part by a grant from the ANR (Corvette project ANR-10-CORD-012).

References

1. Gravano, A., Hirschberg, J.: Turn-taking cues in task-oriented dialogue. Computer Speech & Language 25(3), 601–634 (2011)
2. ter Maat, M.: Response selection and turn-taking for a sensitive artificial listening agent. Ph.D. thesis, University of Twente [Host], Enschede (2011)
3. Ratcliff, R.: A note on modeling accumulation of information when the rate of accumulation change over time. J. of Mathematical Psychology 21, 178–184 (1980)
4. Sacks, H., Schegloff, E.A., Jefferson, G.A.: A simplest systematics for the organisation of turn-taking in conversation. Language 50, 696–735 (1974)
5. Warren, W.H.: The dynamics of perception and action. Psychological Review 113(2), 358–389 (2006)
6. Wilson, M., Wilson, T.P.: An oscillator model of the timing of turn-taking. Psychonomic Bulletin & Review 12(6), 957–968 (2005)